DOCUMENT RESUME

ED 299 921 HE 021 915

AUTHOR Heffernan, James M.; And Others
TITLE Standardized Tests and the Purposes of Assessment.
INSTITUTION American Association for Higher Education, Washington, D.C.
PUB DATE 88
NOTE 27p.; Paper collected as part of the American Association for Higher Education Assessment Forum.
AVAILABLE FROM AAHE Assessment Forum, American Association for Higher Education, One Dupont Circle, Suite 600, Washington, DC 20036.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Educational Assessment; Higher Education; *Outcomes of Education; *Standardized Tests; *Student Evaluation
IDENTIFIERS *AAHE Assessment Forum; *College Outcomes Assessment

ABSTRACT

In December of 1987, the American Association for Higher Education Assessment Forum convened a group of 25 assessment experts to examine the role of standardized tests in assessment. Some of the issues addressed are as follows: what makes standardized testing a particular concern in assessment; what the issues are; how campus assessment leaders think through what is at at stake in the choice they face; and what can be learned about the uses of standardized tests from assessment practitioners. Five parts focus on the following: (1) assumptions about assessment and testing (assessment is not synonymous with standardized testing; much of test bashing is in the air; and college-level assessment is at an early stage of development); (2) the characteristics of standardized tests (they are standardized; commercially available ones are generally designed to elicit comparative information; they provide a contemporaneous expert view of what students should learn; and they have a ready-to-go quality); (3) special issues and concerns (not much is available; these tests are narrow in what they measure; their breadth is a potential problem; they are difficult to connect to specific fixable features of teaching and curriculum; they raise issues of ownership; they raise issues of student involvement; they tend to take on a priviledged but unwarranted status, and they serve the purposes for which they are designed); (4) strategies for effective use (defining purpose; character of information provided; technical properties of the instrument; faculty expectations; student experience; logistics; who will see the information; and where to go for further information); and (5) future developments. (SM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299921

Standardized Tests And The Purposes Of Assessment

by James M. Heffernan,
Pat Hutchings,
and Theodore J. Marchese

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AAHE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 021 915

THE AAHE ASSESSMENT FORUM

AMERICAN ASSOCIATION FOR HIGHER EDUCATION

One Dupont Circle
Suite 600
Washington, D.C. 20036
202/293-6440

AMERICAN ASSOCIATION FOR HIGHER EDUCATION

Board of Directors

Chair
Joseph F. Kauffman
University of Wisconsin
Madison

Chair-Elect
Adele S. Simmons
Hampshire College

Vice Chair
Reatha Clark King
Metropolitan State University

Past Chair
Harriet W. Sheridan
Brown University

Carlos H. Arce
NuStats, Inc.

Estela M. Bensimon
Teachers College
Columbia University

Anne L. Bryant
American Association of
University Women

Donald L. Fruehling
McGraw-Hill, Inc.

Ellen V. Futter
Barnard College

Jerry G. Gaff
Hamline University

Zelda F. Gamson
University of Michigan

Stephen R. Graubard
Daedalus

Joseph Katz
State University of New York
at Stony Brook

Arthur E. Levine
Bradford College

Frank Newman
Education Commission
of the States

Alan Pifer
Carnegie Corporation
of New York

W. Ann Reynolds
The California State
University

Piedad F. Robertson
Miami-Dade Community
College

D. Wayne Silby
Groupware Systems

P. Michael Timpane
Teachers College
Columbia University

President
Russell Edgerton

The AAHE ASSESSMENT FORUM is a three-year project supported by the Fund for the Improvement of Postsecondary Education. It entails three distinct but overlapping activities:

--an annual conference
(the first scheduled for June 14-17, 1987, in Denver)

--commissioned papers
(focused on implementation and other timely assessment concerns; available through the Forum for a small fee)

--information services
(including consultation, referrals, a national directory, and more)

This paper is part of an on-going assessment collection maintained by the Forum. We are pleased to make it more widely available through the ERIC system.

For further information about ASSESSMENT FORUM activities, contact Patricia Hutchings, Director, AAHE ASSESSMENT FORUM, One Dupont Circle, Suite 600, Washington, DC 20036

STANDARDIZED TESTS AND THE PURPOSES OF ASSESSMENT

James M. Heffernan
SUNY College of Environmental
Science and Forestry

Pat Hutchings
AAHE Assessment Forum

and

Theodore J. Marchese
American Association for
Higher Education

Prepared for:

The AAHE Assessment Forum
American Association for
Higher Education

Across the country, campuses are putting together programs to assess student-learning outcomes. Many of these programs are built around standardized tests, an emphasis that raises controversy.

Proponents argue that standardized tests should be at the heart of any assessment, that they provide a reliable accounting not achievable in other ways. Opponents insist that standardized tests are fraught with pitfalls, prone to misuse, and unlikely to provide information of use to students and faculty. In theory, standardized tests serve purposes not easily achieved through other methods. In practice, many campuses have turned to standardized tests as a quick response to an external mandate to do "assessment," only to be disappointed, even angry, at the return. Today, test proponents confront a growing camp of "test bashers."

We argue in this paper that the use of standardized tests to assess student learning is not an either/or proposition. Like other assessment methods, standardized instruments present distinct opportunities and limitations. The useful question is how, and how most effectively, to combine such instruments with other approaches. What are their special features? What do those features tell us about appropriate uses--and misuses? What can be learned from the experience of campuses that have struggled with these questions?

In December, 1987, the AAHE Assessment Forum convened a group of twenty-five assessment experts to examine the role of standardized tests in assessment. Research scholars, state policymakers, representatives of testing organizations, and (most predominantly) campus administrators and faculty assembled at the Belmont Conference Center in Maryland. Discussion was rich with experience and thick with difficult questions, about which there was a considerable range of views. The tone was one of caution about the limits and potential misuses of

standardized tests but also of encouragement to use tests as one of several legitimate methods for assessing student learning and program effectiveness.

Our purpose in this paper is not to force or claim a consensus but to lay out the relevant terrain: What makes standardized testing a particular concern in assessment? What <u>are</u> the issues? How can campus assessment leaders think through what's at stake in the choices they face? What can be learned about the uses of standardized tests from assessment practitioners, particularly those from campuses with more complete programs in place?

Taking the discussions at Belmont as a starting point, this paper attempts to address those questions. The intended audience is campus administrators and faculty who have taken responsibility for designing and implementing assessment programs, efforts that will not only meet an external mandate, if there is one, but lead to real improvements in teaching and learning.

PART I: ASSUMPTIONS ABOUT ASSESSMENT AND TESTING

Several assumptions underlie this paper--assumptions that grounded the Belmont discussions and set an important context for what follows.

First, assessment is not synonymous with standardized testing. This is a point that was frequently--and vigorously--made several years ago, when assessment first moved into the national spotlight; it now bears repeating. Broadly speaking, "assessment" in collegiate settings refers to a range of methods for investigating the phenomena and outcomes of student learning. The usual purpose is that of making judgments about the effects of instruction or curricula, with an aim to their improvement. Because college-level outcomes, and student learning, occur along complex, subtle dimensions, <u>multiple</u> methods are necessary to their understanding.

College-level assessment, then, must be understood not as a single method, test or otherwise; it necessarily calls for methodological variety.

A second assumption follows. A good deal of test bashing is in the air--understandably so, to a point. However, in a comprehensive, multifaceted program of assessment, a reliable standardized instrument does have a role to play. Indeed, all of what are considered the most "mature" assessment programs--even the most student-centered ones, such as that at Alverno College--use standardized tests in one way or another as part of a larger, "multiple-methods" effort.

A short example may be helpful; let us look to the assessment of learning in the major. To understand student learning as an outcome of departmental teaching and curriculum, a range of investigative methods might be drawn upon: group and individual interviews with enrolled and departed students; follow-up reports from employers; a careful, faculty-team reading of student exams and papers; numeric "indicators" of drop-out, graduation, and grad-school acceptance rates; questionnaire data on student perceptions, satisfaction, and effort; visiting faculty and practitioner judgments of senior projects; and, yes, results on an appropriate standardized test--the GRE, for instance, or a licensing exam. Each of these information sources potentially adds insight; findings from several weighed together add up to a fairer, more useful total picture. Within that process, a good, standardized test can make an important, separate contribution to knowledge and belongs in the picture.

Third, and finally, this paper notes as a fact that college-level assessment--visible and pervasive as it has become on the national scene--is nevertheless at an early stage of development. Campuses have little experience with assessment generally; when it comes to standardized tests, inexperience is coupled with limited options. While hundreds of instruments have accumulated over the years for

use with college-student populations, only a few of these were designed with today's assessment needs in mind.

This picture is currently changing, however. As R&D from the testing companies catches up with the new demand for higher-education assessment tools, the array of available standardized instruments will almost certainly expand and change in character. What's most useful then, and what this paper attempts to provide, is not an instrument-by-instrument analysis of what's on the shelf today--"use this," "avoid that"--but more general principles for thinking about and making choices regarding the place of standardized tests in a fuller program of assessment.

PART II: CHARACTERISTICS OF STANDARDIZED TESTS

The concept of standardized tests is a broad and varied one. It's possible to standardize anything from a measure of muscle tone to an exam on quadratic equations. In the simplest terms, "standardized" implies "you can depend on it." A "standardized" test is one that has met various protocols for technical excellence; it may be home-grown or the product of a national firm, employ fixed-response or open-ended questions, be scored by machine or trained readers, and have any number of other characteristics.

In this paper, however, as at the Belmont discussion, what is at issue is standardized tests in their most commonly understood sense: commercially available, machine-scored, multiple-choice instruments that provide comparative data about students. Commonly used examples in assessment today include the ACT-COMP, the ETS Academic Profile, and the College Board's CLEP exams (these three to assess outcomes of general education), the field tests of the Graduate Record Examination (currently being modified for use in assessing outcomes of the major), plus a host of more particular instruments designed to measure a specified

ability (analytical reading or critical thinking) or content domain.

Most of these instruments share characteristics pertinent to their best use in assessment.

1. They are, in fact, "standardized." In lay terms, they have "very good questions." More precisely, if you wanted to ask a question about X, a good test would help you ask it well--with items that have been put through extensive screens for clarity, relevance, parsimony, subject-matter accuracy, appropriate difficulty, and bias. Each item and scale will have met high, public standards of validity and reliability--technical concepts that mean that the test measures what it purports to measure and does so in consistent ways. At Belmont, even critics of standardized tests agreed that the kind of off-the-shelf, standardized instruments with which this paper is concerned represent high levels of technical development, levels theoretically attainable in locally designed instruments but, in practice, expensive and time-consuming to achieve. In short, a standardized test is a test "you can count on."

2. Commercially available standardized tests are almost always designed to elicit comparative information--they answer the question, how well are our students doing compared to X? Because that question can be an important one (especially to external parties), the ability to answer it represents a significant advantage of tests over most other methods.

The "compared with" data presented by tests may take one of two forms. On the one hand the relevant comparison may be normative, with test takers compared to other test takers: Susan compared to Bob, English majors to business majors, this year's seniors to last year's, students at college X to those nationwide. A norm-referenced test typically compares a student or cohort with a national norm group, leading to statements like, "Compared with sophomores

nationally, ours rank in the 65th percentile in mathematics achievement."

On the other hand, the comparison may be to some absolute standard or "criterion." Criterion-referenced tests measure student performance against specified standards and lead to statements like, "Our sophomores as a group know 65% of what is called for" or "Eighty-seven percent have achieved the performance level we expect for graduation."

Both norm- and criterion-referenced tests have their advocates, but it should be noted that they are not mutually exclusive. The Degrees of Reading Power test, for example, was initially developed by the College Board as a criterion-referenced instrument, but now (because of user demand) it provides normative information, too. ETS's Academic Profile, in pilot-test form, was exclusively norm-referenced; an upcoming version will provide criterion-referenced information.

It is worth noting that commercially available, standardized tests are not the only way to answer "compared with" questions. A locally designed instrument given over the years, for example, can track the comparative performance of successive classes; and statistical "indicators" (such as drop-out rates) can be used within or across colleges to compare relative student progress.

3. When looking at student achievement in subject-matter domains, a positive aspect of standardized tests is that they provide a contemporaneous, expert view of what students should learn. Field exams or the GRE, for example, have been written by national committees of scholars; they reflect a disciplinary consensus, of sorts, about what undergraduates in a given field should know. Tests so developed, while far from definitive about what given students in a given department ought to learn, can nevertheless serve as a guard against departmental programs that may be parochial, outdated, or wrongly focused. At the

least, tests designed by disciplinary experts can provide a frame of reference for local introspection, and, for external parties, a standard for looking across the array of nonstandard programs in American higher education.

4. Finally, standardized instruments have a ready-to-go quality. It is, arguably, this convenience factor, more than any psychometric feature, that prompts many campuses to turn first to standardized tests. Typical of politcal time-lines, an assessment mandate from the state calls for something to be done now: standardized tests can be put in place immediately. To the busy dean or department chair there is much to commend a technique that by-passes lengthy, energy-consuming steps of item development and scale construction, validation studies, and complicated data collection and scoring regimes. As with the calorie-counted, nutritionally balanced "Lean Cuisine" meals, it's "heat and eat" with standardized tests. Test materials are easily ordered. They can be administered by a small staff. Scoring is done centrally, by the testing company. Results are returned in a compact, summary form that is simple enough (some would say simplistic) for broad audiences to comprehend--audiences that could (or would) not make their way through, say, lengthy transcripts of student interviews.

Good questions. Comparability. Consenus by experts. Convenience. These are, in themselves, neutral features, a mix of characteristics that distinguishes standardized tests from other methods of assessment. Part IV sets forth strategies for their use that take advantage of these characteristics. Before that, however, it's useful to look more closely at some special issues and concerns raised by the use of standardized tests.

PART III: SPECIAL ISSUES AND CONCERNS

Several concerns about standardized tests were flagged by practitioners at Belmont, concerns that render their use on campus problematic or at least troublesome.

1. A simple, first fact is that, when it comes to standardized tests for assessing student-learning outcomes, not much is available. Unlike entry-level testing for basic skills--the market for which has been visible enough to provoke many years of R&D from the testing companies--few instruments are "ready to go" on the outcomes side. Depending on what is counted, there are three, maybe four instruments marketed for assessing general-education outcomes; two of these are yet at the pilot stage. Similarly, there are few choices related to learning in the major, though, again, more options are on the way. (An apparent exception comes in the area of content-oriented tests; the Mental Measurement Yearbook lists dozens. For many of them, however, quality and appropriateness to assessment purposes are questionable.)

2. Standardized tests are, by design, narrow in what they measure. Psychometrically speaking, the "best" tests --particularly in the ability domain--are also the most limited in focus. A given measure of "critical thinking," for example (five to seven such instruments are in wide use), achieves high levels of reliability and validity by positing a narrow, particular definition of that ability--which may or may not be your definition, but is certain to leave a fair amount of ground uncovered. Lay this circumstance against the breadth and multiplicity of collegiate outcomes and a problem is apparent: even a dozen of the "best" instruments might cover but a fraction of a given college's intended outcomes...another reminder that standardized tests make sense only in the context of a

larger, multifaceted assessment program.

Moreover, the narrowness that ensures psychometric soundness makes many standardized tests uncongenial to faculty, who are often most interested in larger, more integrated outcomes than are likely to be measured by such highly focused methods.

3. Paradoxically, the narrowness of standardized tests is a potential problem, but so, too, is their breadth. Whereas colleges and programs see themselves as having distinct missions, emphases, clienteles, and goals, commercial tests prepared for a broadest-possible market necessarily focus on generic, "common-denominator" outcomes, which may or may not represent the most significant features of a particular program. The problem, then, is one of a mismatch between information provided by the test and information pertinent to actual institutional questions and concerns.

It's a problem that surfaces particularly with broader-scale instruments, such as those designed to assess general education outcomes. The ACT-COMP, for example, is constructed around what many educators see as a group of highly desirable general abilities (using the arts, clarifying values, etc.). However, those abilities may bear little resemblance to actual course and program goals. A content analysis by faculty at the University of Tennessee, Knoxville, indicates that COMP scales match fewer than a third of the goals of that university's general-eduation program. The instrument, in other words, leaves untouched a major part of the UT-K program, just as the character of that program renders big parts of the COMP irrelevant to UT-K. This is by no means a decisive argument against the instrument--no instrument can be expected to match every goal in every college, and even "irrelevant" sections of a test can prompt good discussion. But it reinforces the advice not to overrely on any one approach or to skirt the need for multiple methods.

4. Standardized tests are difficult to connect to specific "fix-able" features of teaching and curriculum. This may be a function either of the character and focus of the test itself or of the ways results are reported.

Many available instruments were simply not designed to provide the kinds of diagnostic information needed for assessment. The SAT and GRE, for example, though both can be found in campus assessment programs, were designed for selection, not for purposes of individual student feedback or program improvement. Similarly, many of today's tests were developed around research questions rather than those asked by college teachers or the assessment community. So a given instrument may report, for instance, about students' "peer independence" or "logical reasoning ability," but what a faculty member in chemistry or a dean of arts and humanities would do with that information is difficult to imagine.

Moreover, because of the way results are reported, even tests that could provide improvement-oriented information may not. Most return data in highly aggregated, summary formats (often, alas, without responses to individual items), so that connecting test results to particular teaching or program features becomes next to impossible.

Finally, it should be said that even tests that focus on real questions about student learning and provide diagnostic results do not, in themselves, ensure improvements. What is true in one way or another of most asessment methods is simply more true of standardized tests: that they provide data, not answers. As one Belmont participant put it, as much time needs to go into "teasing out" the data as to its collection. Steps of interpretation and meaning-making are ever crucial.

5. Standardized tests raise issues of ownership. As off-the-shelf products, they are often--and perhaps inevitably--perceived by faculty as "someone else's answers

to someone else's questions." This reaction may have little to do with the quality of the test itself, but it does highlight crucial political issues.

Belmont participants, many speaking from long experience, were especially adamant that "faculty ownership" is a necessary prerequisite for assessment that will lead to improvement--improvement that must, after all, entail faculty actors. So strong at Belmont were feelings on this point that several veterans argued that commercially available tests were inherently problematic for improvement purposes, and that in all possible cases they should be superseded by local, faculty-developed instruments (these also for the sake of better match with local goals).

6. Standardized tests raise issues of student involvement. Many campuses now find themselves up against the hurdle of students who, having gotten wind that "these things don't count," decline to participate in assessment activities or (even more troublesome) won't take them seriously. This problem is almost inevitably aggravated by standardized tests. To be sat down in a room with several hundred other students for two hours of fill-in-the-circle, multiple-choice items, is a less than rewarding prospect for students. Motivation declines further when students encounter test items about subjects they've never studied (and which therefore leave them feeling "stupid"), and learn that "test security" and scoring procedures limit any possibility of feedback.

Beyond this, the use of standardized tests sends students messages about what we expect of them and what we value. As Alexander Astin points out, norm-referenced tests communicate an ethic of competition, with quality defined as doing better than the next person. Tests that call for right answers selected from among predetermined alternatives communicate to students a low priority on creativity; they denigrate the kinds of learning needed to solve open-ended questions like those encountered in life. Once again, this

is not to say that standardized tests have no role in assessment, but it does suggest considerations beyond the technical and administrative that need to be weighed if tests are to serve rather than subvert larger purposes.

7. Standardized tests tend to take on a privileged but unwarranted status. Thus, while in theory it makes sense to say, as this paper has repeatedly, that standardized tests are best seen as one method among many, in practice their summary, computer-generated test scores and numbers rise to the top, attracting attention in a way that other methods often don't. A case study done in 1988 of state policy toward assessment tells this tale once more: whatever the range of methods required or suggested by a state's mandate, test results inevitably become the focus of attention (and contention).

The danger here, as many colleges have already found, is that the bottom-line scores of a test easily become, in the wrong hands, the object of misuse and abuse. Belmont participants told stories of test scores seized on by the press to draw indefensible conclusions about named colleges and their programs. In several instances, institutional rankings have been concocted by the press from test results (public-disclosure laws allow access to them), the rankings compiled with no heed to institutional mission or clientele or to matters of sample variation or other circumstance. Of course, institutions, too, can be abusers, when they release scores to the press or public on a selective basis to improve their image or political position.

The unwarranted weight assigned to test scores may lead to even more objectionable practices at the level of the individual student, where again, computer quantifications can crowd out "softer," more complete information. The special abuse here is the practice of using a single score as an eligibility screen for specific courses or levels of study. "Everyone knows that a 450

SAT-verbal can't take Chaucer, and that a 500 or below in math shouldn't be allowed into organic chemistry"--shorthand presumptions that far overstep what any test can mean. Indeed, making any important decision about a student (or program) on the basis of a single test score is quite simply indefensible.

8. Standardized tests serve the purposes for which they are designed. Though apparently self-evident, this point was repeatedly noted by Belmont participants troubled by assessment practices which ignore those purposes. Instances were cited of instruments designed for selection being used for placement or for program evaluation, of "gain scores" calculated with tests not meant to measure change.

Particularly dismaying is the possibility that standardized tests of student learning outcomes may be used to evaluate faculty for promotion and tenure. There are several reasons not to link assessment to faculty evaluation, not the least of which is the need for untrammeled faculty involvement (the ownership issue) if improvements are to follow. Even more to the point is that course outcomes are a function of a range of factors (student preparation and effort, for instance), quite beyond the control of any individual faculty member.

A crucial resource for responsble test use is the technical manual which accompanies the instrument. It bears careful reading, and is relevant to many of the concerns discussed above.

PART IV: STRATEGIES FOR EFFECTIVE USE

The characteristics of standardized tests--as we've seen--give them distinct advantages. They have a power to put an issue on the table in a way that's difficult to ignore. They're powerful attention-getters. With these advantages, however, come limitations and potential pitfalls--as we've also seen.

With both the advantages and the dangers of standardized testing in view, Belmont participants discussed a variety of strategies for their effective use, eight of which follow. All assume a larger, multiple-methods approach to assessment. Because they represent ways to think about testing, rather than rules or caveats, each is introduced with a question.

1. What is your purpose, and what methods match that purpose?

An assessment program may of course have several purposes: to evaluate programs or curricula, engage faculty with questions of undergraduate effectiveness, spur or assure student learning, provide for system-level accountability, generate data for marketing or public-relations use--all of these are legitimate reasons to assess. Any one of them may prompt a role for standardized tests. In general, however, existing outcomes-oriented standardized tests, with their emphasis on comparative, highly aggregated data, will be less help for faculty work with individual students than for higher-level, often external decision making.

The point here is not to conjure an absolute rule but to suggest the primacy of <u>purposes</u> over all other considerations in the choice of method. What, after all, do you want to accomplish? Whose questions do you hope to address? These two first-level questions can take you a long way toward choosing appropriate methods.

2. What character of information does the test provide?

This is a question about scoring methods and the qualities of data an instrument will return to you. What scores and subscore reports does the test-maker routinely provide? What further analysis is possible?

Although one of the points of a standardized test is

to provide summary, aggregate data, most practitioners want also the capacity to disaggregate--for example, by program, student background, patterns of course-taking, across scales and by individual item; the greater your power to do so, the more likely you'll be able to answer real questions. You'll want to check, then (in a test's technical manual or with a company representative), about the possibility of conducting such analyses. How much information or customization will the company's initial report back to you provide? Will raw data be available on tape or disk for additional local analysis? What costs and technical requirements will that analysis entail?

3. What are the technical properties of the instrument?

This question speaks to the confidence that may be placed in a test's returns, and therefore to what inferences can and cannot be drawn from them.

A good rule of thumb is to be skeptical about the technical properties of a test. A few standardized tests have a long pedigree of development and use over the years, come with full technical manuals replete with reliability data and validation studies, and have around them active research literatures and user groups--but many do not.

A case in point--one currently on the minds of many assessment practitioners--entails norming practices. Every company wants norms for its tests; not all companies and tests, however, have had the resources behind them to establish and maintain norms. And without sufficient information about norm groups, local comparisions are chancy. What is the size and composition of a given test's norm group? (The norm group may be too broad for comparisons with distinctive local populations.) Why did those in the norm population take the test--as a course requirement or voluntarily? (Different reasons may suggest different levels of motivation.) When were those norms last updated? Marketing materials for some tests claim norms

but are reticent about details. Without such information, claims about "norms" mean little; weak, outdated, inappropriate, or unknown characteristics of a norm group can lead to meaningless or, worse yet, deceptive comparisons involving your students.

Reputable testing organizations stand ready to answer your questions, easy or hard, about their products. Many will, in fact, send a representative to your campus to provide answers and help sort out strategies for best use of their instruments.

4. How well does the test match faculty expectations?

Technical qualities are a necessary but insufficient reason to choose one test over another. A next screen needs to come from faculty. How well do a test's scales match college goals? Does the substance of the test match faculty expectations for what students should know? Is the test's level of difficulty appropriate for our students?

Several campuses represented at Belmont have asked faculty to do a content analysis of tests under consideration or actually to take one or more of them. Debriefing by faculty test-takers can provide insights into the appropriateness of an instrument as well as a substantive way to involve key actors in the development of an assessment program.

5. How will students experience the instrument?

Many campuses find this question should have been asked earlier than it was; it is a factor easy to neglect in a rush to meet mandates or get on a bandwagon.

Asking faculty to analyze or take a test is a step in the right direction; they will, presumably, be in a position to anticipate student boredom, frustration, or the like. Better yet, involve students directly: several institutions have asked student volunteers to take and

compare potential instruments, and to advise on procedures for their administration.

Feedback to students is at issue here, too. Will there be any? Of what character and timing? What steps could make it more meaningful for students?

6. What about logistics? And costs?

Issues of test administration are integral to the student experience; they're also crucial for administrators coping with arrangement-making: Under what conditions should the test be given? Can it be taken in a typical, fifty-minute course slot or does it require a whole morning in the gymnasium? Are there security considerations in scheduling, arranging rooms, assigning proctors? Special equipment needs? What is the turn-around time for scoring? And, most important of all, how can the whole process come before students in ways that motivate their best participation?

The cost of the instruments themselves is easily known; much harder to figure are the indirect costs, in lost opportunties, for instance, or in student/faculty perceptions of what is meant by "assessment." Calculating such costs, and weighing the benefits of a test, is a tricky proposition. A rule of thumb from Belmont practitioners: "It costs more than you think."

7. Who will see the information and what will they do with it?

This question is second in priority only to that of purposes, and surely relates to it. It <u>is</u> a question about purposes: whose questions will the test address? It further asks, what will those parties <u>do</u> with the information when they have it? Unless there's a client for the information, ready and able to act on it, test-giving (or any assessment) will be an empty exercise.

Simultaneously, there need to be, in advance, public decisions about who sees the data and uses it for what purpose...this to keep clear the test's point and guard against mischief.

8. Where can we look for further information and guidance?

Your best source of information always will be experienced colleagues at like institutions. Testing companies can usually provide a list of campuses that have used the instrument you're considering.

In addition, there are now three excellent information sources of a more general nature:

The Assessment Resource Center at the University of Tennessee, Knoxville (supported by the Fund for the Improvement of Postsecondary Education) collects and disseminates information about assessment instruments and methods. For assistance, contact:

Trudy Banta, Director
Gary Pike, Assistant Director
Assessment Resource Center
2046 Terrace Avenue
University of Tennessee, Knoxville
Knoxville, Tennesee 37996-3504
(615) 974-0883

For a more technical treatment of issues in standardized testing (and other assessment methods), plus excellent bibliographies, we recommend a new publication from the U.S. Department of Education's Office of Educational Research and Improvement, Performance and Judgment, a collection of essays edited by Clifford Adelman; $15 (stock #065-000-00342-2) from the Superintendent of Documents, Washington, D.C. 20402.

Finally, for guidelines on ethical issues raised by testing, consult an important document jointly developed by the American Psychological Association, the American

for Measurement in Education, Standards for Educational and Psychological Testing.

PART V: FUTURE DEVELOPMENTS

Assessment is with us to stay, and standardized tests will play a role in many campus assessment programs. The character of that role and how well it's played remain to be seen. How will standardized testing evolve in the next decade? Where will future developments take us?

To answer such questions, it may be helpful to look at the process of test development. As representatives of testing organizations at Belmont were quick to point out, they produce what educators ask for. Producing tests that meet new (and higher) campus expectations, however, can be easier said than done. As one test-maker put it, "The challenge to the testing industry is to find sufficient measures of what educators agree are desirable outcomes, get an instrument developed in a short time, at a reasonable cost, in ways that are psychometrically sound and provide rapid feedback useful to students, faculty and administrators...."

As is obvious, the development of new instruments entails considerable lag-time between user demand and company response. The situation is not unlike Detroit's big-car/small-car dilemma. For decades, test-makers have produced instruments for screening and selection; only in the last few years has a market for outcomes measures become evident. Like the auto industry, then, test-makers face complicated design, retooling, and production adjustments. A number of new--some very promising--developments are now moving forward within the industry, but real progress will take years.

One thing is certain: Useful developments from test-makers depend on clear signals from test-users. Educators leading assessment programs need to understand the

benefits and pitfalls of today's standardized instruments--as we've discussed in this paper. Campuses need to be clear about their own assessment purposes and then insist upon (and help develop) methods which will serve those purposes. Leadership by educators will help bring forward a next, more-promising generation of standardized tests for assessment.

BELMONT PARTICIPANTS

Clifford Adelman
Senior Associate
Office of Research
U.S. Department of Education

Richard T. Alpert
Division Director of Operations
National Evaluation Systems, Inc.

Trudy Banta
Professor
Learning Research Center
University of Tennessee, Knoxville

Nancy Beck
Director
College and University Programs
Educational Testing Service

Richard L. Burns
Senior Development Leader
Educational Testing Service

Karen Carey
Professor of Educational Policy Studies and Evaluation
University of Kentucky

Tom Carroll
Acting Deputy Director
Fund for the Improvement of Postsecondary Education

Donald Carstensen
Educational Service Division
American College Testing Program

Constance Cook
Program Officer
Fund for the Improvement of Postsecondary Education

Wade C. Curry
Dean of Arts and Sciences
Trenton State College

Russell Edgerton
President
American Association for Higher Education

Willard Enteman
Provost
Rhode Island College

Peter Ewell
Senior Associate
National Center for Higher Education Management Systems

Donald Farmer
Academic Dean
King's College

Arnold Gelfman
Director of Planning and Special Projects
Brookdale Community College

James Heffernan
Vice President for Student Affairs and Educational Services
SUNY College of Environmental Science and Forestry

Milton Hillery
Assistant Vice President for Assessment Services
American College Testing Program

Pat Hutchings
Director
AAHE Assessment Forum

Larry Jones
Associate Director
Institutional Research and Planning
University of Georgia

Kathleen Kies
Deputy Executive Director
Minnesota Higher Education Coordinating Board

Richard Light
Professor of Education
Harvard University

Ted Marchese
Vice President
American Association for Higher Education

Marcia Mentkowski
Director
Office of Research and Evaluation
Alverno College

Jason Millman
Professor of Educational Research Methodology
Cornell University

E. Thomas Moran
Assistant Vice President for Academic Affairs
SUNY College at Plattsburgh

Scott Olsen
Associate Professor of Education
Northeast Missouri State University

Robert Thoeny
Executive Director
Washington Higher Education Coordinating Board

Orlando Toro
College Level Examination Program
The College Board

ADDITIONAL RESOURCES
AAHE ASSESSMENT FORUM

The following resources are available for purchase from the AAHE Assessment Forum:

1. **Resource Packet I: Five Papers** **$15.00**
   --"Assessment, Accountability, and Improvement: Managing the Contradiction," P. Ewell
   --"Assessment and Outcomes Measurement: A View from the States," C. Boyer, P. Ewell, J. Finney, and J. Mingle
   --"The External Examiner Approach to Assessment," B. Fong
   --"Six Stories: Implementing Successful Assessment," P. Hutchings
   --"Thinking About Assessment: Perspectives for Presidents and Chief Academic Officers," E. El-Khawas and J. Rossmann

2. **Resource Packet II: Six Papers** **$25.00**
   --"Acting Out State-Mandated Assessment: Evidence from Five States," C. Boyer and P. Ewell
   --"Assessing Student Learning in Light of How Students Learn," J. Novak and D. Ridley
   --"Faculty Voices on Assessment: Expanding the Conversation," P. Hutchings and E. Reuben
   --"Feedback in the Classroom: Making Assessment Matter," K. Patricia Cross
   --"Standardized Tests and the Purposes of Assessment," J. Heffernan
   --"An Update on Assessment," (AAHE Bulletin, December, 1987), P. Hutchings and T. Marchese

3. **Three Presentations--1987:** **$8.00**
   from the 2nd National Conference on Assessment in Higher Education
   --Lee S. Shulman -- "Assessing Content and Process: Challenges for the New Assessments"
   --Virginia B. Smith -- "In the Eye of the Beholder: Perspectives on Quality"
   --Donald M. Stewart -- "The Ethics of Assessment"

4. **Three Presentations--1988:** **$10.00**
   from the 3rd National Conference on Assessment in Higher Education
   --Alexander Astin -- "Assessment and Human Values: Confessions of a Reformed Number Cruncher"
   --Linda Darling-Hammond -- "Assessment and Incentives: The Medium is the Message"
   --Robert H. McCabe -- "The Assessment Movement: What Next? Who Cares?"

5. **Assessment Programs and Projects: A Directory** **$10.00**
   Concise descriptions of thirty assessment projects being implemented on campuses across the country. Edited by Jacqueline Paskow.

To order items indicated above, contact: Patricia Hutchings, Director, AAHE Assessment Forum, One Dupont Circle, Suite 600, Washington, DC 20036; 202/293-6440. Orders under $25 must be prepaid. Allow four weeks for delivery. **MAKE CHECKS PAYABLE TO AAHE ASSESSMENT FORUM.**